Movie and Script Alignment without Dialogues

Supervisor at INRIA : Rémi Ronfard, IMAGINE team, remi.ronfard@inria.fr
Supervisor at TECHNICOLOR : Patrick Perez, Technicolor R & D, Rennes.

Context

The alignment of a movie and its script provides insight into the movie making process and can be used to build annotated data sets useful in many areas, including actor and action recognition. A common approach, first described by Ronfard and Thuong [1] consists in separating dialogues from action descriptions, and aligning the dialogues with the subtitles in the movie [2,3,4]. Cour et al have taken this approach even further by using the alignments as a starting point for also aligning scene breaks and for naming actors in the movie [5]. One problem with such approaches is that they break down in actions scenes with little or no dialogue. In such cases, the alignment needs to be performed based on visual cues such as object and actor appearances, actor movements and actions, which are challenging and difficult problems [6].

Objectives

In this master's thesis, we will explore methods used in machine translation [7,8] and adapt them to the case of movie/script alignment by discovering frequently occurring visual cues and motion patterns from the video that can be mapped to actor names and motion verbs in the script [9,10].

The candidate will have to answer the following questions : (1) which elements of the script and the movie can be matched to each other manually ? (2) which subset of those matches can be obtained automatically, using state of the art computer vision ? and (3) assuming that an optimal alignment between the text and the movie is one that maximizes the number of correct matches, how can such an optimal alignment be obtained in finite time using dynamic programming techniques ?

For evaluation, scenes from the silent movie "the artist" will be annotated and aligned with its (English) script, both manually and automatically.

The Master thesis will be supervised by Patrick Perez at Technicolor in Rennes and Remi Ronfard at INRIA in Grenoble, and is expected to lead to a PHD thesis on the same topic at Technicolor in 2013-2015.

For this thesis, we are looking for candidates with a strong background in mathematics and scientific programming (matlab), with an interest in computer vision and computational linguistics.

References

[1] Remi Ronfard, Tien Tran Thuong. A framework for aligning and indexing movies with their script. Multimedia and Expo, 2003. ICME'03.

[2] Marcin Marszałek, Ivan Laptev, Cordelia Schmid. Actions in Context. CVPR 2009.

[3] Ivan Laptev Marcin Marszałek Cordelia Schmid Benjamin Rozenfeld. Learning realistic human actions from movies. CVPR 2008.

[4] Pramod Sankar K., C. V. Jawahar and Andrew Zisserman. Subtitle-free Movie to Script Alignment Proceedings of the 20th British Machine Vision Conference (BMVC 2009)

[5] Timothee Cour, Chris Jordan, Eleni Miltsakaki, Ben Taskar. Movie/Script: Alignment and Parsing of Video and Text Transcription. ECCV 2008.

[6] Abhinav Gupta, Praveen Srinivasan, Jianbo Shi and Larry S. Davis, Understanding Videos, Constructing Plots: Learning a Visually Grounded Storyline Model from Annotated Videos, In CVPR 2009.

[7] P. Duygulu, Kobus Barnard, J. F. G. de Freitas, and David A. Forsyth. 2002. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. In Proceedings of the 7th European Conference on Computer Vision-Part IV (ECCV '02)

[8] Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. Comput. Linguist. 19, 2 (June 1993), 263-311.

[9] Daniel Weinland, Remi Ronfard, Edmond Boyer. Automatic Discovery of Action Taxonomies from Multiple Views. CVPR 2006.

[10] I. Junejo, E. Dexter, I. Laptev, P. Pérez. View-Independent action recognition from temporal self-similarities. IEEE Trans. Pattern Anal. Machine Intell. 33(1), 2011.